

When I say ... computerised adaptive testing

Citation for published version (APA):

Collares, C. F., & Cecilio-Fernandes, D. (2019). When I say ... computerised adaptive testing. *Medical Education*, 53(2), 115-116. <https://doi.org/10.1111/medu.13648>

Document status and date:

Published: 01/02/2019

DOI:

[10.1111/medu.13648](https://doi.org/10.1111/medu.13648)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

When I say . . . computerised adaptive testing

Carlos Fernando Collares^{1,2}  & Dario Cecilio-Fernandes^{3,4} 

Traditional paper-based knowledge tests require student cohorts to answer the same sets of questions. These fixed tests do not consider students' levels of knowledge. For students with low levels of knowledge, a particular paper-based test might be too difficult, whereas for students with high levels of knowledge, the same test is likely to be too easy. The mismatch between students' knowledge levels and test difficulty may lead to lower levels of reliability, and the items may not discriminate low and high levels of performance. Given that videos and sounds could increase the professional authenticity of assessment, another limitation of paper-based tests may be that they do not allow the use of multimedia.

Computerised adaptive testing (CAT) is a form of computer-based assessment in which an algorithm dynamically selects the next item based on the student's previous answers.¹ Consequently, each student receives an individually customised test, tailored to his or her level of knowledge. A basic requirement of CAT is a calibrated item bank that allows the algorithm to choose the item difficulty of the next question based on the answers provided. This calibration is usually based on item response theory and estimates item difficulty and student ability using the same scale. To properly calibrate the questions, assumptions of unidimensionality and local independence should be met; this, in turn, provides evidence of the validity of the internal

structure of CAT. In the educational context, however, validity based on test content is more important than any validity evidence based on the internal structure of CAT. In CAT, validity based on test content can be achieved by constraining the content to the desired percentage of the blueprint, covering all the necessary content.¹ By matching item difficulty with student ability, CAT may reduce the length of the test by 50% while keeping or even increasing reliability. Reliability estimates tend to be more homogeneous regardless of students' knowledge levels. As CAT provides short, reliable and valid knowledge tests, it may be used as a strategy for the assessment of learning.

One of the most neglected aspects of CAT is its alignment with modern learning theories, such as constructivism. Vygotsky described the zone of proximal development as the frontier of one's knowledge, where optimal learning is expected to occur. By matching test difficulty with student ability, we dare to say that CAT brings the zone of proximal development to the assessment realm,² although it can also be argued that CAT provides a rough and imperfect approximation to the zone of proximal development. In addition, as CAT may minimise the occurrence of both cognitive under- and overload, it may be aligned with cognitive load theory. With more reliable scores, students are less likely to have improper mastery experiences as a result of unreliable scores, which could negatively

¹Department of Educational Development and Research, School of Health Professions Education, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, the Netherlands

²European Board of Medical Assessors, Maastricht, the Netherlands

³Centre for Education Development and Research in Health Professions (CEDAR), University Medical Centre Groningen, University of Groningen, Groningen, the Netherlands

⁴Werkgroep Interuniversitaire Voortgangstoets Geneeskunde, the Netherlands

Correspondence: Carlos Fernando Collares, Department of Educational Development and Research, School of Health Professions Education, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht 6229 ER, the Netherlands. Tel: 00 31 43 388 5798; E-mail: c.collares@maastrichtuniversity.nl

affect their self-efficacy.³ Students are eager for accurate, interactive and even immediate feedback, which is one of the main features of CAT. When used within an assessment programme, CAT can further maximise the educational utility of knowledge testing. Therefore, it is possible to say that CAT is in line with the concept of assessment *for learning*.

Students seem to accept CAT very well. In our experience, after switching to CAT, students say that they become more motivated to study because they perceive their scores to be more accurate representations of their knowledge levels. Surprisingly, students who failed have also shared similar perceptions with us. It could be argued that student surveys and student evaluations of CAT do not provide sufficient evidence to support such perceptions. However, recent evidence suggests that CAT may also have positive impacts on students' achievement, motivation, engagement and subjective test experiences in comparison with paper-based testing.⁴ The results also suggested positive effects on the performances of older students, who are typically seen as less motivated and engaged. Future research should include the impact of CAT on health professions education, a frontier yet to be explored.

Considering its improved psychometric properties, positive educational impact and high level of acceptability, CAT may optimise the utility of assessment tools, as proposed by van der Vleuten.⁵ Although CAT has been used for decades in the assessment *of learning*, we acknowledge that further

research regarding the use of CAT in medical education is required. More modern CAT approaches include multi-stage adaptive testing and cognitive diagnostic adaptive testing, which may be more suitable approaches in medical education as they can better deal with measures that require the testing of multiple dimensions. Based on recent developments and evidence, an increase in the use of CAT in medical education under an assessment *for learning* paradigm can be expected. Further research to explore the didactical use of computerised adaptive tests as learning moments is recommended.

REFERENCES

- 1 Chang HH. Psychometrics behind computerized adaptive testing. *Psychometrika* 2015;**80** (1):1–20.
- 2 Allal L, Ducrey GP. Assessment of – or in – the zone of proximal development. *Learn Instr* 2000;**10** (2):137–52.
- 3 Cilliers FJ, Schuwirth LW, Adendorff HJ, Herman N, van der Vleuten CP. The mechanism of impact of summative assessment on medical students' learning. *Adv Health Sci Educ Theory Pract* 2010;**15** (5):695–715.
- 4 Martin AJ, Lazendic G. Computer-adaptive testing: implications for students' achievement, motivation, engagement, and subjective test experience. *J Educ Psychol* 2018;**110** (1):27–45.
- 5 van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ Theory Pract* 1996;**1** (1):41–67.

Received 13 March 2018; editorial comments to authors 26 April 2018; accepted for publication 30 May 2018